

Formation Big Data, Méthodes et solutions pratiques pour l'analyse des données

Présentation

Cette formation intermédiaire propose une approche pragmatique du Big Data appliqué à l'analyse des données. Sur trois jours, les participants apprennent à analyser des besoins, mettre en œuvre des pipelines d'ingestion et de préparation, configurer des traitements distribués (batch/stream) et produire des visualisations et indicateurs exploitables en contexte entreprise.

Durée : 21,00 heures (3 jours)

Tarif INTRA : [Nous consulter](#)

Objectifs de la formation

- Analyser des besoins analytiques et définir un périmètre Big Data adapté
- Mettre en œuvre l'ingestion et la préparation à l'échelle (formats colonnes, partitionnement)
- Configurer des traitements distribués (batch/stream) et orchestrer des workflows
- Mettre en œuvre des visualisations et tableaux de bord orientés décision
- Analyser la performance et la qualité des données pour fiabiliser les usages

Prérequis

- Connaissances de base en SQL
- Notions de manipulation de données (CSV, JSON)
- Familiarité avec un langage de script (Python de préférence)

Public

- Data analysts, data engineers et développeurs évoluant vers la data à l'échelle
- Équipes BI/IT souhaitant industrialiser l'analyse de données volumineuses

Programme de la formation



Jour 1 – Ingestion, stockage et préparation à l'échelle

Session du matin :

- Panorama Big Data pour l'analyse : cas d'usage, limites et bonnes pratiques
- Modélisation des zones de données (bronze/silver/gold), data lake et lakehouse
- Formats et optimisation : Parquet/ORC, partitionnement, gestion des schémas

Session de l'après-midi :

- Ingestion multi-sources (fichiers, API) et validation de schémas
- Préparation des données : nettoyage, jointures, agrégations, enrichissement
- Contrôles de qualité : règles, métriques et traçabilité

TP / Exercice :

- Construire un pipeline d'ingestion et de préparation (CSV/JSON → format colonne) avec contrôle qualité (livrables : scripts, diagramme de flux, rapport qualité)

Points clés & takeaways :

- Choisir formats et stratégies d'ingestion adaptés au volume
- Mettre en place une chaîne de préparation reproductible et contrôlée

Jour 2 – Traitements distribués et orchestration

Session du matin :

- Traitements batch et streaming : principes, usages et arbitrages
- Performance : partitions, parallélisme, coût vs latence
- Tests et observabilité : logs, métriques, profiling de jobs

Session de l'après-midi :

- Orchestration (Airflow, Prefect) : DAG, dépendances, retries, SLA
- Mise en œuvre d'un job distribué (Spark ou équivalent)
- Reprise sur incident, idempotence et fiabilité

TP / Exercice :

- Développer un workflow orchestré pour alimenter un datamart (batch + micro-batch) avec suivi de performance (livrables : DAG, code, métriques)

Points clés & takeaways :

- Exécuter et fiabiliser des traitements à l'échelle
- Mettre en place une observabilité utile à l'exploitation

Jour 3 – Visualisation, KPIs et restitution décisionnelle**Session du matin :**

- Conception d'indicateurs : définitions, granularité, périodicité
- Bonnes pratiques de visualisation et storytelling
- Publication et gouvernance : data contracts, sécurité (aperçu)

Session de l'après-midi :

- Construction d'un dashboard de bout en bout (KPI, tendances, alertes)
- Documentation et checklist d'exploitation (qualité, coûts, rafraîchissement)
- Revue finale : limites, améliorations et feuille de route

TP / Exercice :

- Créer un dashboard alimenté par le pipeline (KPI, filtres, alertes) avec note d'exploitation

Points clés & takeaways :

- Produire des restitutions utiles et actionnables
- Relier pipeline, qualité des données et décision métier

Organisation

Formateur

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes. Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétences.

Moyens pédagogiques et techniques

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation

En amont de la formation *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

Tout au long de la formation

- Évaluations continues réalisées par le formateur travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

À l'issue de la formation *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation
- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

Accessibilité

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.